



University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): PA Thwaites

Article Title: Causal identifiability via Chain Event Graphs

Year of publication: 2011

Link to published article:

<http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2011/paper11-16>

Publisher statement: None

Causal identifiability via Chain Event Graphs

Peter Thwaites

Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom

Abstract

We present the Chain Event Graph (CEG) as a complementary graphical model to the Causal Bayesian Network for the representation and analysis of causally manipulated asymmetric problems. CEG analogues of Pearl's Back Door and Front Door theorems are presented, applicable to the class of *singular* manipulations, which includes both Pearl's basic *Do* intervention and the class of functional manipulations possible on Bayesian Networks. These theorems are shown to be more flexible than their Bayesian Network counterparts, both in the types of manipulation to which they can be applied, and in the nature of the conditioning sets which can be used.

Keywords: Back Door theorem, Bayesian Network, causal identifiability, causal manipulation, Chain Event Graph, conditional independence, Front Door theorem

1. Introduction

In this paper we consider cause and effect through the analysis of controlled models. The standard apparatus for such an approach is the Causal Bayesian Network (CBN) [4, 8, 9, 18]. As noted in [22], CBNs are ideal for problems which have a natural product space structure, but need adaptation for problems which do not. It is this latter type of problem that we are primarily concerned with here.

Context-specific variants of Bayesian Networks (BNs) have been developed for tackling asymmetric problems [1, 7, 12, 14]. These are still rather awkward for the representation and analysis of problems whose future development at any specific point depends on the particular history of the

Email address: `Peter.Thwaites@warwick.ac.uk` (Peter Thwaites)

April 14, 2011

problem upto that point, and the values of a particular set of covariates at that point. Their use is similarly circumscribed in problems where there may be no possible outcomes of some variables given certain histories or values of covariates.

There have of course been many recent advances in CBN theory (see for example [2, 3, 5, 10, 23, 24]), some of which have made the causal analysis of asymmetric problems simpler. However even with these advances the available graphical representations for such problems and the types of manipulation an analyst can consider are still limited. Similarly the available analytical techniques are often rather crude. It was argued in [22] that *causes* are more naturally expressed as *events* rather than the values of some random variable. The Chain Event Graph (CEG) introduced in [16] provides an ideal graphical **representation** given this argument. It is also a sensible representation for the **analysis** of manipulations to events. Moreover, as shown in [22], use of the CEG makes available a richer class of possible manipulations than is generally the case with CBNs.

The collection of techniques available for use with CEG-based causal analysis is already sufficient for tackling most problems, if not yet as large as that available for BN-based analysis. A Back Door theorem for CEGs analogous to Pearl's [8, 9] Back Door theorem for BNs was introduced in [22]. Here we present a much more general version of this as well as two versions of a Front Door theorem, the second of which allows considerably more flexibility than the analogous BN version [8, 9]. We anticipate that future work will replicate for CEGs the work done in [3, 23, 24] which provides necessary and sufficient conditions for causal identifiability in BNs.

As the CEG is a comparatively new structure, there have been minor modifications since [16], and indeed since [22]. These are detailed in the next section. We believe these changes improve the CEG by making it less messy, and also by turning it into a genuine directed acyclic graph (DAG), which latter allows us to utilise the many results proven for this graph type.

In Section 2 we define the CEG and manipulated CEG. Section 3 develops the Back Door theorem and the idea of *singular manipulations*. A Front Door theorem, a generalisation of Pearl's [8, 9] theorem for BNs, is then introduced in Section 4, and Section 5 provides a discussion of possible directions for future research.

2. Definitions and notation

In this section we give a brief definition of a CEG. This has been modified slightly since [16] and [22]. We also provide some notation that will be used throughout the paper. We then turn our attention to what it means when we manipulate a CEG to an event, and present a definition of a manipulated CEG.

The CEG is a function of an event tree [15], retaining those features of the tree which allow for the transparent representation of asymmetric problems. They are a significant extension to trees since they express within their topology the entire conditional independence structure of the problems which they have been created to represent [20].

An event tree \mathcal{T} is a directed tree with vertex set $V(\mathcal{T})$ and edge set $E(\mathcal{T})$. The root-to-leaf paths $\{\lambda\}$ of \mathcal{T} form the atoms of the event space. Events measurable with respect to this space are unions of these atoms.

Each non-leaf vertex $v \in V(\mathcal{T})$ labels a random variable $X(v)$ whose state space $\mathbb{X}(v)$ can be identified with the set of directed edges $e(v, v') \in E(\mathcal{T})$ emanating from v . For each $X(v)$ we let

$$\Pi(v) \equiv \{\pi_e(v' \mid v) \mid e(v, v') \in \mathbb{X}(v)\}$$

where $\pi_e(v' \mid v) \equiv P(X(v) = e(v, v'))$ are called the *primitive probabilities* of the tree; and

$$\Pi(\mathcal{T}) \equiv \{\Pi(v)\}_{v \in V(\mathcal{T})}$$

Definition 1. (Coloured tree) For an event tree \mathcal{T} with vertex set $V(\mathcal{T})$ and edge set $E(\mathcal{T})$

1. Two non-leaf vertices $v^1, v^2 \in V(\mathcal{T})$ are in the same stage u if there is a bijection $\psi(v^1, v^2)$ between $\mathbb{X}(v^1)$ and $\mathbb{X}(v^2)$ such that if $\psi : e(v^1, v^{1'}) \mapsto e(v^2, v^{2'})$ then $\pi_e(v^{1'} \mid v^1) = \pi_e(v^{2'} \mid v^2)$. The edges $e(v^1, v^{1'})$ and $e(v^2, v^{2'})$ have the same colour if v^1 and v^2 are in the same stage, and $e(v^1, v^{1'})$ maps to $e(v^2, v^{2'})$ under this bijection.
2. Two vertices $v^1, v^2 \in V(\mathcal{T})$ are in the same position w if for each subpath emanating from v^1 , the ordered sequence of colours is the same as that for some subpath emanating from v^2 .

The set of stages of the tree is labelled $L(\mathcal{T})$, and the set of positions is labelled $K(\mathcal{T})$.

In the definition of the CEG below we have removed the undirected edges from previous definitions, and introduced colouring of nodes. We believe this makes the CEG easier to read, and it also allows us to utilise the extensive theory relating to DAGs.

Definition 2. (Chain Event Graph)

The Chain Event Graph $\mathcal{C}(\mathcal{T})$ is the coloured DAG with vertex set $V(\mathcal{C})$ and edge set $E(\mathcal{C})$ defined by:

1. $V(\mathcal{C}) \equiv K(\mathcal{T}) \cup \{w_\infty\}$.
2. (a) *For $w, w' \in V(\mathcal{C}) \setminus \{w_\infty\}$, there exists a directed edge $e(w, w') \in E(\mathcal{C})$ iff there are vertices $v, v' \in V(\mathcal{T})$ such that $v \in w \in K(\mathcal{T})$, $v' \in w' \in K(\mathcal{T})$ and there is an edge $e(v, v') \in E(\mathcal{T})$.*
 (b) *For $w \in V(\mathcal{C}) \setminus \{w_\infty\}$, there exists a directed edge $e(w, w_\infty) \in E(\mathcal{C})$ iff there is a non-leaf vertex $v \in V(\mathcal{T})$ and a leaf vertex $v' \in V(\mathcal{T})$ such that $v \in w \in K(\mathcal{T})$ and there is an edge $e(v, v') \in E(\mathcal{T})$.*
3. *If $v^1 \in w^1 \in K(\mathcal{T})$, $v^2 \in w^2 \in K(\mathcal{T})$ and v^1, v^2 are members of the same stage $u \in L(\mathcal{T})$, then we say that w^1, w^2 are in the same stage u , and assign the same colour to these positions. We label the set of stages of \mathcal{C} by $L(\mathcal{C})$.*
4. *If $v \in w \in K(\mathcal{T})$, $v' \in w' \in K(\mathcal{T})$ and there is an edge $e(v, v') \in E(\mathcal{T})$, then the edge $e(w, w') \in E(\mathcal{C})$ has the same colour as the edge $e(v, v')$.*

The root-to-sink paths $\{\lambda\}$ of \mathcal{C} form the atoms of the event space of \mathcal{C} . Events measurable with respect to this space are unions of these atoms.

Each stage $u \in L(\mathcal{C})$ labels a random variable $X(u)$ whose state space $\mathbb{X}(u)$ can be identified with the set of directed edges $e(w, w') \in E(\mathcal{C})$ emanating from any $w \in u$.

Example 1. CEG construction

We illustrate the construction of a CEG through a fault diagnosis example, which for illustrative convenience uses only binary variables.

- A machine utilises two components $C1$ and $C2$. Whether $C2$ is functioning properly or is faulty is independent of whether $C1$ is functioning properly or is faulty.

Table 1: Context for Example 1

Descriptor	Edges
$C1$ functioning properly	$e(v_0, v_1)$
$C1$ faulty	$e(v_0, v_2)$
$C2$ functioning properly	$e(v_1, v_3), e(v_2, v_5)$
$C2$ faulty	$e(v_1, v_4), e(v_2, v_6)$
$C3$ on & functioning properly	$e(v_4, v_7), e(v_5, v_9), e(v_6, v_{11})$
$C3$ on & faulty	$e(v_4, v_8), e(v_5, v_{10}), e(v_6, v_{12})$
$C2$ replaced by $C2'$, $C2'$ functioning properly	$e(v_8, v_{13}), e(v_{12}, v_{17})$
$C2$ replaced by $C2'$, $C2'$ faulty	$e(v_8, v_{14}), e(v_{12}, v_{18})$
$C1$ replaced by $C1'$, $C1'$ functioning properly	$e(v_{10}, v_{15})$
$C1$ replaced by $C1'$, $C1'$ faulty	$e(v_{10}, v_{16})$

same position (since they root isomorphic coloured subtrees). The vertices v_8 & v_{12} are in the same stage and the same position.

The CEG in Figure 2 illustrates the ideas of Definition 2. The vertices v_4 & v_6 from the tree have been merged into one position w_4 representing $C2$ faulty. The positions w_1 & w_2 are in the same stage (indicated by the colouring of the nodes) since whether or not $C2$ functions properly is independent of whether or not $C1$ functions properly. The positions w_3 & w_4 are in the same stage as they represent $C3$ switching on given $C1$ faulty and $C2$ either functioning properly or not.

The following notation will be used throughout the remainder of the paper. Recall that an atom λ is a $w_0 \rightarrow w_\infty$ path in \mathcal{C} . The set of atoms is denoted Ω . We write $w \prec w'$ when the position w precedes the position w' on a $w_0 \rightarrow w_\infty$ path. We call w a *parent* of w' if there exists an edge $e(w, w') \in E(\mathcal{C})$.

Events are denoted Λ . $\Lambda(w)$ is the event which is the union of all $w_0 \rightarrow w_\infty$ paths passing through the position w , and $\Lambda(e(w, w'))$ is the union of all paths passing through the edge $e(w, w')$.

We can now define the *primitive probabilities* of the CEG: $\pi_e(w' | w)$ is the probability of the edge $e(w, w')$; and for each $u \in L(\mathcal{C})$ and random variable $X(u)$ we let

$$\Pi(u) \equiv \{\pi_e(w' | w) \mid w \in u\}$$

and

$$\Pi(C) \equiv \{\Pi(u)\}_{u \in L(C)}$$

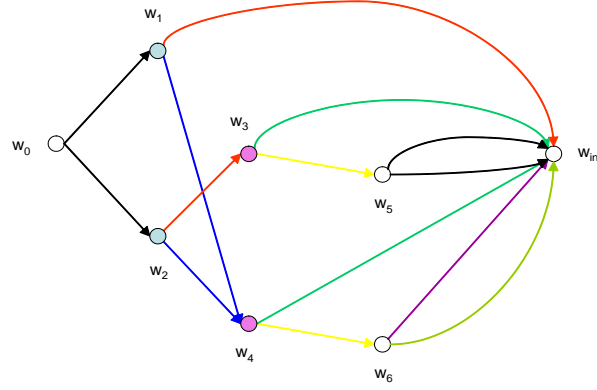


Figure 2: Chain Event Graph for Example 1

Note that if we label the probability of the event Λ by $\pi(\Lambda)$ then $\pi_e(w' | w) \equiv \pi(\Lambda(e(w, w')) | \Lambda(w))$.

A subpath of a root-to-sink path is denoted $\mu(w, w'')$, where w and w'' indicate the start and end positions of the subpath. $\Lambda(\mu(w, w''))$ is the event which is the union of all paths utilising the subpath $\mu(w, w'')$. $\pi_\mu(w'' | w) \equiv \pi(\Lambda(\mu(w, w'')) | \Lambda(w))$ is the probability of the subpath $\mu(w, w'')$.

Before moving on to manipulated CEGs we present a very useful Lemma, proofs of which appear in [19] and [20].

Lemma 1. *For a CEG \mathcal{C} and positions $w_1, w_2, w_3 \in V(\mathcal{C})$ such that $w_1 \prec w_2 \prec w_3$*

$$\pi(\Lambda(w_3) | \Lambda(w_1), \Lambda(w_2)) = \pi(\Lambda(w_3) | \Lambda(w_2))$$

The result can be extended so that the positions w_1 & w_2 can each be replaced by edges, and the position w_3 can be replaced by a collection of positions and/or edges.

Essentially this tells us that being at a position (w_3) or edge (or collection of positions or edges), given that we have been at an earlier position (w_2) or edge, is independent of the path taken to that position or edge. This result is used in the proof of Theorem 1.

2.1. Manipulated CEGs

Anything that we observe about a system or do to a system will change the topology of a graphical representation of that system. In [21] we considered how the topology of a CEG is altered when we observe an event Λ . Here we investigate how the topology of a CEG is altered when we manipulate to an event Λ . As the following definitions suggest, the process of updating our beliefs following a manipulation is very similar to that which happens following the observation of an event. Note that the use of trees in causal analysis has a respectable history, featuring in for example [13, 15, 18].

For the purposes of this paper we assume that the CEG is valid (in that it satisfies the conditions of [22] Definition 3) for any manipulation we choose to make. A detailed discussion of what makes a CEG valid for a causal manipulation can be found in [22] Sections 3.1 and 3.2.

The type of events we consider in this paper are *intrinsic* events [19, 20] (called \mathcal{C} -compatible events in [21]). An intrinsic event Λ is one where every atom of Λ is a $w_0 \rightarrow w_\infty$ path of a subgraph of \mathcal{C} , and every $w_0 \rightarrow w_\infty$ path in this subgraph is an atom of Λ .

Definition 3. (Manipulated CEG) For a CEG $\mathcal{C}(V, E)$ and intrinsic event Λ , let $\hat{\mathcal{C}}^\Lambda$ (the CEG manipulated to the event Λ) be the subgraph of \mathcal{C} with

- (a) $V(\hat{\mathcal{C}}^\Lambda) \subset V(\mathcal{C})$ contains precisely those positions which lie on a $w_0 \rightarrow w_\infty$ path $\lambda \in \Lambda$
- (b) $E(\hat{\mathcal{C}}^\Lambda) \subset E(\mathcal{C})$ contains precisely those edges which lie on a $w_0 \rightarrow w_\infty$ path $\lambda \in \Lambda$
- (c) For $w_1, w_2 \in V(\hat{\mathcal{C}}^\Lambda)$, and $e(w_1, w_2) \in E(\hat{\mathcal{C}}^\Lambda)$, the edge $e(w_1, w_2)$ has probability uniquely assigned by the definition of the manipulation to Λ (by say Definition 6 or [22] Definition 3)
- (d) If $w_1, w_2 \in V(\hat{\mathcal{C}}^\Lambda)$ are in the same stage in $\hat{\mathcal{C}}^\Lambda$ then these positions and their emanating edges are coloured in $\hat{\mathcal{C}}^\Lambda$.

Probabilities in $\hat{\mathcal{C}}^\Lambda$ are denoted $\hat{\pi}^\Lambda$. For completeness we also define a conditioned CEG (see also [21]).

Definition 4. (Conditioned CEG) For a CEG $\mathcal{C}(V, E)$ and intrinsic event Λ , let \mathcal{C}^Λ (the CEG conditioned on the event Λ) be the subgraph of \mathcal{C} with $V(\mathcal{C}^\Lambda)$, $E(\mathcal{C}^\Lambda)$ defined and coloured analogously with $V(\hat{\mathcal{C}}^\Lambda)$, $E(\hat{\mathcal{C}}^\Lambda)$ in Definition 3, and

(c) For $w_1, w_2 \in V(\mathcal{C}^\Lambda)$, and $e(w_1, w_2) \in E(\mathcal{C}^\Lambda)$, the edge $e(w_1, w_2)$ has probability

$$\pi_e^\Lambda(w_2 \mid w_1) = \frac{\sum_{\lambda \in \Lambda} \pi(\lambda, \Lambda(e(w_1, w_2)))}{\sum_{\lambda \in \Lambda} \pi(\lambda, \Lambda(w_1))}$$

where π^Λ indicates a probability in \mathcal{C}^Λ and π a probability in \mathcal{C} .

3. The Back Door theorem

Pearl's [8, 9] Back Door theorem for BNs provides a condensed version of the full manipulated probability expression. So when a manipulation is impossible or unethical in practice, or its effects difficult or impossible to observe, an analyst may still be able to estimate the probabilities of the theoretically possible effects of this manipulation.

Since 1995 there has been considerable effort put in to finding conditions for causal identifiability on BNs [3, 10, 11, 23, 24] – that is conditions for when the effects of a manipulation can be estimated from a subset of variables observed in the idle system. CEG-based causal theory is unsurprisingly not so far advanced. The Back Door theorem for CEGs introduced in [22] is however already more flexible than its counterpart for BNs, as we demonstrate here.

Pearl's Back Door theorem for BNs states that under certain conditions on sets of variables X, Y, Z , we can (using the notation of [6]) write the probability of observing $Y = y$ following a manipulation of X to x as

$$p(y \parallel x) = \sum_z p(y \mid x, z) p(z)$$

As already implied, this expression requires the analyst to observe only the idle (or unmanipulated) system and condition on these observations. By careful choice of the set Z we may be able to calculate or estimate $p(y \parallel x)$ without conditioning on the full set of measurement variables.

One rather useful aspect of the theorem is that the conditions can be expressed graphically (that is, on the BN of the problem).

The Back Door theorem for CEGs introduced in [22] is valid for a larger collection of types of manipulation than are possible with a BN, and since it refers to manipulation to events rather than of variables, it is more consistent with our experience of what a manipulation actually involves. As with the BN version of the theorem, we reduce the complexity of the general manipulated probability expression, as well as reducing or avoiding identifiability problems associated with it.

So consider a manipulation to the event Λ_x . Suppose we wish to find the probability of (observing) an event Λ_y given that the manipulation to Λ_x has been enacted – that is we wish to produce an expression for $\pi(\Lambda_y \parallel \Lambda_x)$. This is equal to the probability of the event Λ_y on the CEG $\hat{\mathcal{C}}^{\Lambda_x}$, which is the sum of the probabilities of the $w_0 \rightarrow w_\infty$ paths in $\hat{\mathcal{C}}^{\Lambda_x}$ which are consistent with the event Λ_y :

$$\pi(\Lambda_y \parallel \Lambda_x) = \hat{\pi}^{\Lambda_x}(\Lambda_y)$$

Note also that $\pi(\Lambda_y \mid \Lambda_x) = \pi^{\Lambda_x}(\Lambda_y)$.

Consider a partition of the atomic events ($w_0 \rightarrow w_\infty$ paths in \mathcal{C}) $\{\Lambda_z\}$. Then

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \hat{\pi}^{\Lambda_x}\left(\bigcup_z \Lambda_z, \Lambda_y\right) = \sum_z \hat{\pi}^{\Lambda_x}(\Lambda_z, \Lambda_y)$$

since the events $\{\Lambda_z\}$ form a partition of Ω

$$= \sum_z \hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda_z) \hat{\pi}^{\Lambda_x}(\Lambda_z)$$

Definition 5. (Back Door partition) *The partition $\{\Lambda_z\}$ forms a Back Door partition of Ω if*

- (A) $\hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda_z) = \pi(\Lambda_y \mid \Lambda_x, \Lambda_z) \quad (= \pi^{\Lambda_x}(\Lambda_y \mid \Lambda_z))$
- (B) $\hat{\pi}^{\Lambda_x}(\Lambda_z) = \pi(\Lambda_z) \quad \text{for all } \Lambda_z \in \{\Lambda_z\}$

If these conditions are satisfied then

$$\pi(\Lambda_y \parallel \Lambda_x) = \hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z) \pi(\Lambda_z)$$

The sets of variables Z in the BN-based Back and Front Door theorems are called *blocking* sets because they *block* certain paths between X and Y in the BN. In a much less transparent way Z also blocks the effect on Y of other problem variables so that the manipulated probability expression $p(y \parallel x)$ can be condensed. The Back Door theorem for CEGs works in an altogether less mysterious way. The blocking set becomes a partition of the $w_0 \rightarrow w_\infty$ paths of the CEG into sets $\{\Lambda_z\}$ which allow us to replace probabilities evaluated on the manipulated graph by ones evaluated on the idle CEG. As with the BN version, if we choose $\{\Lambda_z\}$ carefully, we can calculate or estimate $\pi(\Lambda_y \parallel \Lambda_x)$ from a partially observed idle system.

3.1. Singular manipulations

Definition 6. (Singular manipulation) A manipulation to Λ of a CEG \mathcal{C} is called singular if there exist sets $W \subset V(\mathcal{C})$, $E_\Lambda \subset E(\mathcal{C})$ such that

- (i) the elements of W partition Ω (ie. every $w_0 \rightarrow w_\infty$ path in \mathcal{C} passes through precisely one $w \in W$),
- (ii) for each $w \in W$, there exists precisely one emanating edge $e(w, w')$ which is an element of E_Λ ,
- (iii) Λ is the union of precisely those $w_0 \rightarrow w_\infty$ paths that pass through some $e(w, w') \in E_\Lambda$,
- (iv) all edge probabilities in $\hat{\mathcal{C}}^\Lambda$ are equal to the corresponding edge probabilities in \mathcal{C} , except that $\hat{\pi}_e^\Lambda(w' | w) = 1$ for $w \in W$, $e(w, w') \in E_\Lambda$.

Essentially, a singular manipulation is one where every $w_0 \rightarrow w_\infty$ path passes through *one* of a collection of positions, and the manipulation imposes a probability of 1 on *one* edge emanating from each of these positions.

All $Do\ X = x$ and functional manipulations (but not all stochastic manipulations) of BNs are singular manipulations, but the set of singular manipulations is much larger than this.

Note that if the manipulation to an event Λ is singular then edge probabilities in $\hat{\mathcal{C}}^\Lambda$ upstream and downstream of the manipulation remain as in the idle CEG \mathcal{C} . If we were to **condition** on this event Λ , then edge-probabilities in \mathcal{C}^Λ downstream of the observation would remain as in the idle CEG, but edge-probabilities upstream would change in accordance with Definition 4 (c).

3.2. A Back Door theorem for singular manipulations

As we also consider effect events (Λ_y) and conditioning sets (Λ_z), we distinguish our manipulation event Λ by adding a suffix to give Λ_x . We also relabel the set W as W_X , the positions within W_X as w_X , and the edges of Defn. 6 (ii) as $e(w_X, w'_X)$.

As the set of positions in W partitions Ω , we can consider a random variable X , defined on Ω , which takes values labelled by the emanating edges of w_X (for each w_X) with probabilities dependent on the history of the problem up to that position w_X .

The manipulation to Λ_x assigns a probability of 1 to one of the values of X at each w_X , dependent on the history of the problem up to that position w_X (ie. a functional manipulation). So Λ_x is of the form

$$\Lambda_x \equiv \bigcup_{w_X \in W_X} \Lambda(e(w_X, w'_X))$$

We define an effect variable Y in exactly the same way as we have defined X . So we have a set of positions W_Y (downstream of the set W_X) which partitions Ω (ie. every $w_0 \rightarrow w_\infty$ path in \mathcal{C} passes through one of the positions in W_Y). Then Λ_y consists of all paths that passing through some $w_Y \in W_Y$, utilise some prespecified edge emanating from that w_Y . So

$$\Lambda_y \equiv \bigcup_{w_Y \in W_Y} \Lambda(e(w_Y, w'_Y))$$

If we look at the conditions for Pearl's Back Door theorem on BNs, we see that both conditions can be re-expressed as conditional independence statements (see for example [2]). Pearl's condition that Z (the Back Door *blocking set*) must block all Back Door paths from X to Y can be expressed as

$$Y \perp\!\!\!\perp Q(X) \mid (X, Z)$$

where $Q(X)$ indicates the variable parents of X . Pearl's condition that Z must contain no descendants of X can be expressed as

$$Z \perp\!\!\!\perp X \mid Q(X)$$

Note that we are here ignoring the possibility that $Z \equiv Q(X)$. We return to this case in section 3.4.

We have already replaced $X = x$ by Λ_x , and $Y = y$ by Λ_y . We now replace $Z = z$ by Λ_z , and noting that positions store the *relevant* history of a problem upto that point, $Q(X) = q(x)$ by $\Lambda(w_X)$.

Substituting into $Z \perp\!\!\!\perp X \mid Q(X)$ we get

$$\begin{aligned} \pi(\Lambda_z \mid \Lambda(w_{X(1)})) &= \pi(\Lambda_z \mid \Lambda(w_{X(1)}), \Lambda_x) \\ &= \pi(\Lambda_z \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda(e(w_X, w'_X))) \\ &= \pi(\Lambda_z \mid \Lambda(e(w_{X(1)}, w'_{X(1)}))) \end{aligned} \tag{3.1}$$

Substituting into $Y \perp\!\!\!\perp Q(X) \mid (X, Z)$ we get

$$\begin{aligned}
\pi(\Lambda_y \mid \Lambda_x, \Lambda_z) &= \pi(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x, \Lambda_z) \\
&= \pi(\Lambda_y \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda(e(w_X, w'_X)), \Lambda_z) \\
&= \pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda_z) \tag{3.2}
\end{aligned}$$

and

$$\pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda_z) = \pi(\Lambda_y \mid \Lambda(e(w_{X(2)}, w'_{X(2)})), \Lambda_z)$$

Theorem 1. (Back Door theorem) *With W_X , W_Y , Λ_x , Λ_y defined as above, and $\{\Lambda_z\}$ a partition of the atomic events, then $\{\Lambda_z\}$ is a Back Door partition if conditions (3.1) and (3.2) hold for all elements of $\{\Lambda_z\}$, $w_X \in W_X$.*

A proof of this theorem appears in the appendix.

Note that these conditions are on the graph \mathcal{C} . They can therefore, like Pearl's conditions, be checked on an unmanipulated graph (a representation of the idle system).

3.3. Checking the conditions for the Back Door theorem

Pearl's conditions for his Back Door theorem can be checked directly on the topology of the BN. For the CEG condition (3.1) requires that for each element Λ_z of the Back Door partition (which could be of the form $\Lambda(w)$, $\Lambda(e)$, a union of such events, or some totally different type of event), and each position $w_X \in W_X$, the probability of Λ_z conditioned on $\Lambda(w_X)$ is the same as that of Λ_z conditioned on the event $\Lambda(e(w_X, w'_X))$ where $e(w_X, w'_X)$ is the singular edge emanating from w_X which remains in the manipulated graph. It is however not immediately apparent how to check condition (3.2) as simply.

Clearly the exact nature of $\{\Lambda_z\}$ is something that we can control. As suggested above we can choose sets of $w_0 \rightarrow w_\infty$ paths to belong to any individual element Λ_z in many different ways. In [22] we let our blocking set consist of events associated with positions upstream of the manipulation. As is the case with BNs, blocking sets cannot be associated with variables that are descendants of the manipulated variable(s), but they don't need to be ancestors. So CEG blocking sets can also be created using positions (or

edges) downstream of the manipulation. Indeed, if the events which we wish to condition on correspond to values of a variable which has not been observed at the time of the manipulation, and if our CEG has been constructed in an extensive form order then our blocking set **must** use positions or edges downstream of the manipulation. In this paper we offer a generalisation of the Back Door theorem of [22], but do not intend to duplicate the results therein. We therefore concentrate in this section on blocking sets downstream of the manipulation. Between the two papers we cover all possible locations for blocking sets of this form.

For the remainder of Section 3 we use partitions where each Λ_z is an event associated with a collection of positions. Replicating this work for events associated with edges is straightforward. So we let each Λ_z be a union of smaller events of the form $\Lambda(w_z^i)$ for some set of positions $\{w_z^i\}$, where this set is a subset of W_z , which is in turn a set of positions downstream of W_X and upstream of W_Y , partitioning Ω . We can of course make the partitions coarser or finer as we see fit.

So a typical element of $\{\Lambda_z\}$ will be of the form

$$\Lambda_z = \bigcup_{i \in A} \Lambda(w_z^i)$$

for some set A , $w_z^i \in W_z$.

The process we describe may appear complicated, but as illustrated in Example 2 it is in fact comparatively straightforward.

As both $\{\Lambda_z\}$ and W_z are partitions of Ω , we can specify that

$$\Lambda(w_z^i) \cap \Lambda(w_z^j) = \phi$$

for $i, j \in A \cup B \cup \dots \cup N$, where N is the number of elements we have specified for $\{\Lambda_z\}$.

Now, whereas all elements of W_z exist in \mathcal{C} , not all will exist in $\hat{\mathcal{C}}^{\Lambda_x}$. As we have control over the nature and coarseness of our partition, we can let N equal the number of elements of W_z which exist in $\hat{\mathcal{C}}^{\Lambda_x}$, and construct each Λ_z so that it contains only one w_z which exists in $\hat{\mathcal{C}}^{\Lambda_x}$. For each Λ_z , call this position w_z^1 . So, however many positions $\{w_z^i\}$ correspond to each element of $\{\Lambda_z\}$, there will be only N positions $\{w_z^1\}$ that exist in $\hat{\mathcal{C}}^{\Lambda_x}$.

The complete set $W_z = \{w_z^i\}_{i \in A \cup \dots \cup N}$ partitions Ω . So

$$\pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda_z) = \pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda(w'_{X(1)}), \Lambda_z)$$

since $\Lambda(e(w_{X(1)}, w'_{X(1)})) \subset \Lambda(w'_{X(1)})$ in \mathcal{C}

$$\begin{aligned} &= \frac{\pi(\Lambda_z, \Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda(w'_{X(1)}))}{\pi(\Lambda_z \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda(w'_{X(1)}))} \\ &= \frac{\pi(\Lambda_z, \Lambda_y \mid \Lambda(w'_{X(1)}))}{\pi(\Lambda_z \mid \Lambda(w'_{X(1)}))} \end{aligned}$$

using the forms specified for Λ_z , Λ_y ; W_z and W_Y being downstream of W_X ; and the result of Lemma 1. Hence

$$\pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda_z) = \pi(\Lambda_y \mid \Lambda(w'_{X(1)}), \Lambda_z) \quad (3.3)$$

But any path-segment in \mathcal{C} starting at $w'_{X(1)}$ remains in $\hat{\mathcal{C}}^{\Lambda_x}$, and we know that $\{w_z^i\}_{i \geq 2}$ do **not** exist in $\hat{\mathcal{C}}^{\Lambda_x}$, so there are no path-segments joining $w'_{X(1)}$ to w_z^i (for $i \geq 2$) in $\hat{\mathcal{C}}^{\Lambda_x}$, and hence no path-segments joining $w'_{X(1)}$ to w_z^i (for $i \geq 2$) in \mathcal{C} . Therefore

$$\Lambda(w'_{X(1)}) \cap \Lambda(w_z^i) = \phi \quad \text{for } i \geq 2$$

and

$$\Lambda(w'_{X(1)}) \cap \Lambda_z = \Lambda(w'_{X(1)}) \cap \Lambda(w_z^1)$$

so expression (3.3) becomes

$$\begin{aligned} \pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda_z) &= \pi(\Lambda_y \mid \Lambda(w'_{X(1)}), \Lambda(w_z^1)) \\ &= \pi(\Lambda_y \mid \Lambda(w_z^1)) \end{aligned}$$

using the form specified for Λ_y ; the fact that W_z is downstream of W_X ; and the result of Lemma 1. Hence

$$\pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w'_{X(1)})), \Lambda_z) = \pi(\Lambda_y \mid \Lambda(e(w_{X(2)}, w'_{X(2)})), \Lambda_z)$$

as required for condition (3.2).

So if we choose each Λ_z to be of the form described above, where for each Λ_z only w_z^1 exists in $\hat{\mathcal{C}}^{\Lambda_x}$, then this is sufficient for condition (3.2) to be satisfied. We now have two conditions which can be checked simply on the topology of the idle CEG.

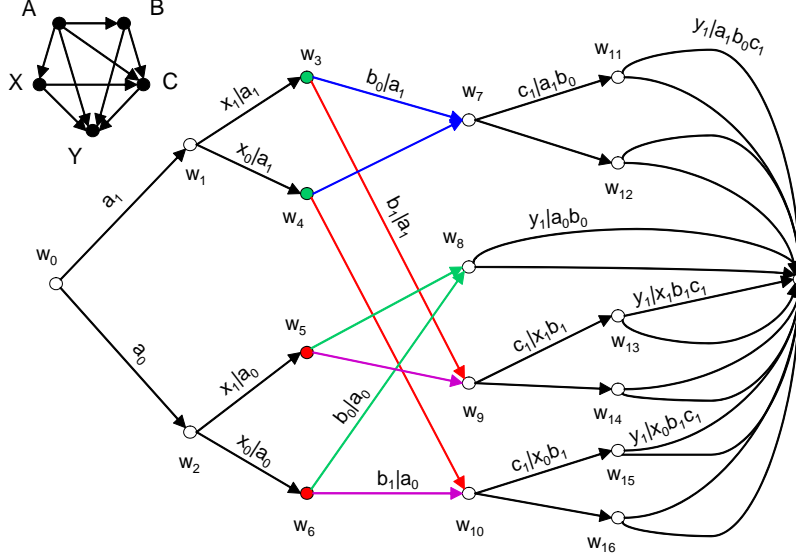


Figure 3: BN and CEG \mathcal{C} for Example 2

Example 2. Using the Back Door theorem

We illustrate the use of our Back Door theorem through a medical example. As with Example 1 we use binary variables for illustrative convenience.

Our interest is in a condition which can manifest itself in one of two forms ($C = 1$ or 2). Individuals who will as adults develop the condition (in either of its forms) display either symptom S_A before the age of ten, or S_B in their late teens, or both. Whether or not an individual displays S_A is labelled by a variable A , and whether or not they display S_B by a variable B . In both cases the variable takes the value 1 if the symptom is displayed, and the value 0 if it is not. There is a treatment T available which has some efficacy if given in an individual's early teens. Being treated is labelled $X = 1$, and not treated $X = 0$. Dying before the age of fifty is labelled $Y = 1$, and dying at fifty or older $Y = 2$.

The relationships between the variables A, X, B, C and Y are described below, and are portrayed by the CEG in Figure 3, where for convenience edges are labelled a_0 for $A = 0$ etc.

Symptom S_A is often missed by doctors, but if it is detected an individual is more likely to be given treatment T . We therefore do not know the

distributions of A , $X \mid A = 0$ or $X \mid A = 1$. We do know however that $X \perp\!\!\!\perp A$.

Evidence from previous studies indicates that

- whether or not an individual displays symptom S_B depends only on whether or not they displayed symptom S_A ($B \perp\!\!\!\perp X \mid A$),
- displaying either symptom means that an individual **will** develop the condition in one of its two forms,
- for individuals displaying S_A but not S_B , developing the condition in form 1 does not depend on whether or not they had treatment T ($C \perp\!\!\!\perp X \mid A = 1, B = 0$). Also, how long they live depends only on which form of the condition they develop ($Y \perp\!\!\!\perp X \mid A = 1, B = 0, C$),
- for individuals displaying S_B , developing the condition in form 1 does not depend on whether or not they displayed S_A , irrespective of whether they were treated or not ($C \perp\!\!\!\perp A \mid X, B = 1$). Also, how long they live depends on whether or not they were treated and on which form of the condition they develop ($Y \perp\!\!\!\perp A \mid X, B = 1, C$).

If we were to attempt to portray the problem via a BN it would look like the one in Figure 3. Without considerable annotation the BN cannot express the context-specific conditional independence structure illustrated by the CEG.

We are interested in the effects on life expectancy (the variable Y) if we were to treat everybody in the population in their early teens. So we consider the singular manipulation to Λ_x equivalent to $Do\ X = 1$, and calculate the probability $\pi(\Lambda_y \parallel \Lambda_x) \equiv P(Y = 1 \parallel X = 1)$. The CEG satisfies the conditions that every path passes through a position from $W_X = \{w_1, w_2\}$ and a position from $W_Y = \{w_8, w_{11}, w_{12}, \dots, w_{16}\}$. Also, every position in W_X has an outgoing edge labelled x_1 ($X = 1$), and every position in W_Y has an outgoing edge labelled y_1 ($Y = 1$).

Clearly A is a required variable in any Back Door blocking set Z based on the BN representation of the problem. But from above we do not know the distribution of A or of any joint distribution involving A . Can we use our Back Door theorem for CEGs to find an identifiable expression not involving A ?

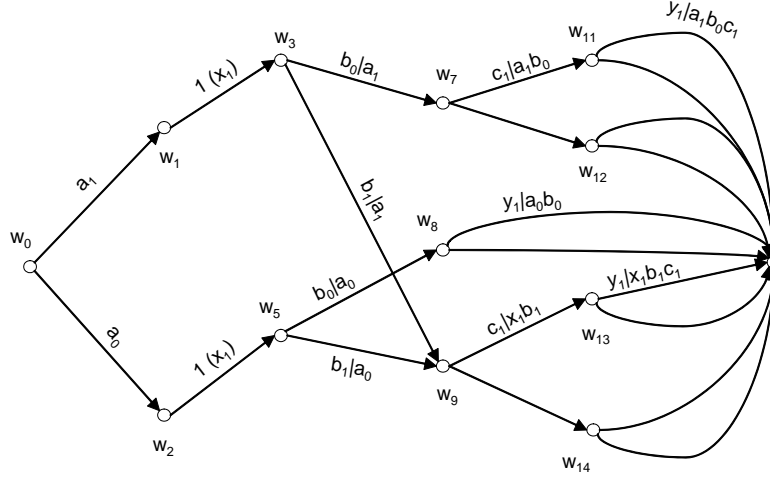


Figure 4: Manipulated CEG $\hat{\mathcal{C}}^{\Lambda_x}$ for Example 2

In these situations we generally have a lot of flexibility in determining our blocking set (Z), and some experimentation may be needed before we find the ideal allocation. Here we are considering Λ_z of the form $\bigcup \Lambda(w)$. The choice of positions will depend on what we can observe, and may be heavily influenced by observation costs. Note that the connection between these constraints and our choice of positions can be very subtle – in this example we clearly cannot estimate $P(A = 1, B = 0, C = 1)$, but we can still include the position w_{11} in our blocking set. Here we simply imagine that these constraints and our experimentation have produced a blocking set of positions W_z , lying between W_X and W_Y , comprising $\{w_8, w_9, w_{11}, w_{12}, w_{15}, w_{16}\}$. The CEG $\hat{\mathcal{C}}^{\Lambda_x}$ is given in Figure 4.

Here $\{e(w_X, w'_X)\} = \{e(w_1, w_3), e(w_2, w_5)\}$, and we combine our $\{w_z\}$ to produce $\{\Lambda_z\}$ as follows

$$\{\Lambda(w_8), \Lambda(w_{11}), \Lambda(w_{12}), [\Lambda(w_9) \cup \Lambda(w_{15}) \cup \Lambda(w_{16})]\}$$

Note that (i) $\{\Lambda_z\}$ forms a partition of Ω , (ii) each Λ_z is of the form $\bigcup \Lambda(w)$, and (iii) three of the Λ_z are *singleton* $\Lambda(w_z)$ where w_z appears in $\hat{\mathcal{C}}^{\Lambda_x}$, and the fourth Λ_z is the union of three $\Lambda(w_z)$ only one of which w_z is present

in $\hat{\mathcal{C}}^{\Lambda_x}$. So condition (3.2) is satisfied.

It is straightforward to show that our $\{\Lambda_z\}$ satisfy condition (3.1). Using the CEG \mathcal{C} we get, for $w_X = w_1$ that

$$\begin{aligned}\pi(\Lambda(w_{11}) \mid \Lambda(w_1)) &= [p(x_1 \mid a_1) + p(x_0 \mid a_1)] p(b_0 \mid a_1) p(c_1 \mid a_1 b_0) \\ &= p(b_0 c_1 \mid a_1) \\ \pi(\Lambda(w_{11}) \mid \Lambda(e(w_1, w_3))) &= p(b_0 \mid a_1) p(c_1 \mid a_1 b_0) = p(b_0 c_1 \mid a_1)\end{aligned}$$

and similarly for the expression involving w_{12} .

The position w_8 is not downstream of w_1 .

$$\begin{aligned}\pi(\Lambda(w_9) \cup \Lambda(w_{15}) \cup \Lambda(w_{16}) \mid \Lambda(w_1)) &= p(x_1 \mid a_1) p(b_1 \mid a_1) \\ &\quad + p(x_0 \mid a_1) p(b_1 \mid a_1) p(c_1 \mid x_0 b_1) \\ &\quad + p(x_0 \mid a_1) p(b_1 \mid a_1) p(c_2 \mid x_0 b_1) \\ &= p(b_1 \mid a_1) \\ \pi(\Lambda(w_9) \cup \Lambda(w_{15}) \cup \Lambda(w_{16}) \mid \Lambda(e(w_1, w_3))) &= \pi(\Lambda(w_9) \mid \Lambda(e(w_1, w_3))) \\ &= p(b_1 \mid a_1)\end{aligned}$$

A similar procedure for $w_X = w_2$ confirms that $\{\Lambda_z\}$ satisfy condition (3.1), and so that $\{\Lambda_z\}$ is a Back Door partition of Ω . Our manipulated probability expression

$$p(y_1 \mid x_1) = \pi(\Lambda_y \mid \Lambda_x) = \hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z) \pi(\Lambda_z)$$

is evaluated on \mathcal{C} , and simplifies to

$$p(b_0) p(y_1 \mid b_0) + p(b_1) p(y_1 \mid x_1 b_1)$$

So we need only know the distribution of B (the incidence of symptom S_B), and the conditional distributions of Y (life expectancy) on the events $B = 0$ (S_B not displayed) and $X = 1, B = 1$ (treated and S_B displayed). This expression does not involve A (the incidence of S_A), and interestingly neither does it involve C (which form the condition takes). It does however involve B , which would be impossible if we used the BN from Figure 3 for this model, as B does not block all Back Door paths from X to Y .

To summarise, the procedure is

- Produce $\{\Lambda_z\}$ as prescribed above, and check that it satisfies our Back Door condition (3.1).

- Substitute probabilities from \mathcal{C} into our Back Door expression and simplify.

This example gives an insight into how to choose the component Λ_z of our partition. If we can find w_z such that $\Lambda(w_z)$ satisfies

$$\pi(\Lambda(w_z) \mid \Lambda(e(w_X, w'_X))) = \pi(\Lambda(w_z) \mid \Lambda(w_X)) \quad \forall w_X \in W_X$$

then we can make $\Lambda(w_z)$ a Λ_z .

Other Λ_z are produced by combining **one** position w_z that exists in $\hat{\mathcal{C}}^{\Lambda_x}$ with other positions $\{w_z\}$ that disappear when we create $\hat{\mathcal{C}}^{\Lambda_x}$, in such a way that the union of their associated events satisfies the Back Door condition (3.1) for all $w_X \in W_X$.

3.4. Using W_X to create a blocking set

Blocking sets using positions upstream of the set W_X were considered in [22]. Here we look at using the set W_X itself to create our blocking set. This has a direct analogy with analysis on BNs, where it is always possible to replace Pearl's set Z by the set $Q(X)$ to give a revised Back Door expression

$$p(y \mid x) = \sum_{q(x)} p(y \mid x, q(x)) p(q(x))$$

This blocking set $Z = Q(X)$ is not derived from the conditions $Z \perp\!\!\!\perp X \mid Q(X)$ and $Y \perp\!\!\!\perp Q(X) \mid (X, Z)$, and similarly our Back Door partition $\{\Lambda_z\}$ here is not derived from conditions (3.1) and (3.2). Recalling the analogy between $Q(X) = q(x)$ for BNs and $\Lambda(w_X)$ for CEGs suggests we look at a partition $\{\Lambda_z\}$ where each Λ_z is of the form

$$\Lambda_z = \bigcup_{i \in A} \Lambda(w_{X(i)})$$

for some set A , where $\Lambda(w_X)$ for each $w_X \in W_X$ is an element of some Λ_z .

The analogy between $Q(X) = q(x)$ for BNs and $\Lambda(w_X)$ for CEGs is not perfect. It is shown in [22] that a better analogy for parents in a BN is a set of *stages*, rather than positions. So here we make a further stipulation about the sets $\{w_{X(i)}\}_{i \in A}$, and state that each Λ_z is of the form

$$\Lambda_z = \bigcup_{w_X \in u_X} \Lambda(w_X) = \Lambda(u_X)$$

for some u_X , where each u_X is a stage, and the set $\{u_X\}$ form a partition of Ω . We also require that for each $w_X \in u_X$, the edges $e(w_X, w'_X)$ carry the same label. These labels can differ for different stages.

This is not actually an onerous restriction, as the set of manipulations we can consider clearly still contains all basic *Do* interventions on BNs and all functional *Do* interventions where the argument of the function is (a subset of) the parent set of the manipulated variable. In fact we can argue that this set contains **all** functional *Do* interventions of a BN: If a manipulation is functional in that the value we manipulate X to depends on the value taken by another variable W , then essentially we have a *decision* problem and the BN representation of the system becomes an Influence Diagram (ID) representation with X as a decision node. Clearly the value of W must be known before X is manipulated, so in this ID representation there must be an edge from W to X (see for example [17]) and so W is a parent of X . Hence we argue that for **all** functional *Do* interventions on BNs the argument of the function is (a subset of) the parent set of the manipulated variable.

In order to demonstrate that the set $\{\Lambda_z\}$ is a Back Door partition we need the result of the following Lemma, a proof of which appears in [19].

Lemma 2. *For a CEG \mathcal{C} , $w_X \in V(\mathcal{C})$, $w_X \in u_X \in L(\mathcal{C})$, and Λ_x defined as in section 3.2*

$$\pi(\Lambda_x \mid \Lambda(w_X)) = \pi(\Lambda_x \mid \Lambda(u_X))$$

This seemingly innocuous result tells us that the probability of leaving a stage by an edge carrying a particular label is the same as that of leaving any of its component positions by an edge carrying this label.

The equality holds if the edges $e(w_X, w'_X)$ label the **same** value of X for each $w_X \in u_X$. This is the case for all basic *Do* interventions and all functional *Do* interventions as described above.

Using the proof of Theorem 1 (in the appendix) we can write

$$\begin{aligned}
\hat{\pi}^{\Lambda_x}(\Lambda_y) &= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \pi(\Lambda_y \mid \Lambda(e(w_X, w'_X))) \\
&= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_x) \\
&= \sum_{u_X} \sum_{w_X \in u_X} \pi(\Lambda(w_X)) \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_x) \\
&= \sum_{u_X} \sum_{w_X \in u_X} \left[\frac{\pi(\Lambda(w_X), \Lambda_x, \Lambda_y)}{\pi(\Lambda_x \mid \Lambda(w_X))} \right] \\
&= \sum_{u_X} \left[\frac{\sum_{w_X \in u_X} \pi(\Lambda(w_X), \Lambda_x, \Lambda_y)}{\pi(\Lambda_x \mid \Lambda(u_X))} \right] = \sum_{u_X} \left[\frac{\pi(\Lambda(u_X), \Lambda_x, \Lambda_y)}{\pi(\Lambda_x \mid \Lambda(u_X))} \right] \\
&= \sum_{u_X} \pi(\Lambda_y \mid \Lambda(u_X), \Lambda_x) \pi(\Lambda(u_X))
\end{aligned}$$

So we can use the set W_X to create a blocking set if we insist that each Λ_z is $\Lambda(u_X)$ for some stage u_X , and that the edges $e(w_X, w'_X)$ label the same value of X for each w_X in any u_X .

Our Back Door theorem for CEGs makes causal analysis with them more flexible than with BNs. Firstly they are ideal for the analysis of asymmetric controlled models such as treatment regimes. Secondly we can analyse the effects of asymmetric manipulations, a task which is not necessarily straightforward on a BN, particularly if both the manipulated variable and the value this variable takes are dependent on the values of other variables. These functional manipulations often require the addition of edges to BN representations which can cause difficulties for an analyst trying to find suitable blocking sets.

Lastly we can use asymmetric blocking sets with CEGs. Recall that a good Back Door expression allows the analyst to estimate probabilities of effects from a partially observed system, so this flexibility in our choice of partition set is very useful when some of the events in the system are unobservable or have large observational costs. Standard causal analysis with BNs requires one to be able to calculate or estimate $p(z)$ and $p(y \mid x, z)$ for all values z of the blocking set of variables Z . This is not necessary with CEGs – our blocking sets do not need to correspond to any fixed subset of the measurement random variables that define a BN. We have also seen

that we can use the CEG version of the Back Door theorem in cases where it would be impossible to use the BN version, as the model does not obey the conditions specified by Pearl. Note that it would not be at all difficult for us to create a Back Door partition which for example consisted of some positions $\{w_z\}$ downstream of the manipulation together with some stages $\{u_X\}$ coincident with the manipulation.

4. A Front Door theorem for CEGs

Pearl's Front Door theorem [8, 9] can be used in cases where the Back Door theorem conditions do not hold or where the events needing to be observed for the Back Door theorem have too large an observational cost. Like the Back Door theorem, the Front Door theorem allows one to reduce the complexity of the general manipulated probability expression used with BNs, and can allow one to sidestep identifiability problems associated with it.

Pearl's Front Door theorem states that under certain conditions on sets of variables X, Y, Z , we can write

$$p(y \parallel x) = \sum_z p(z \mid x) \sum_{x'} p(y \mid x', z) p(x')$$

an expression whose value can be estimated from a partially observed idle system.

The expression for the Front Door theorem is more complex than that for the Back Door theorem, and this imposes greater restrictions on the types of manipulation we can consider and also **initially** on the nature of our blocking sets. So we confine ourselves here to singular manipulations and note that as our initial expression will be directly analogous to that for BNs, we will need to sum over some variable corresponding to Pearl's X . Hence we need to produce a partition of Ω , of which Λ_x is **one** element. Realistically this means confining ourselves to start with to manipulations directly analogous to Pearl's *Do* $X = x$ (for some criterion variable X), and consider positions $\{w_X\}$ which each have the same number of emanating edges and where these edges carry the same labels for each w_X (ie. each w_X has an emanating edge labelled x_j for j in some set J).

Note that even for fairly regular problems depictable by BNs there may be histories or parental configurations of a variable X for which the probability

of a particular outcome is zero. Although normally we do not draw zero-probability edges in a CEG, in this case it is advisable to do so, if only for the edges emanating from those positions associated with the variable X .

In section 4.2 we see that we can relax these conditions considerably, and that there is a version of the Front Door theorem for CEGs which is significantly more flexible than Pearl's Front Door theorem for BNs.

Pearl quotes three conditions for using the Front Door theorem, but these can actually be reduced to two conditional independence conditions

$$Y \perp\!\!\!\perp X \mid (Z, Q(X)) \quad \text{and} \quad Z \perp\!\!\!\perp Q(X) \mid X$$

Using the same approach as for the Back Door theorem, we can suggest appropriate CEG versions of these conditions. We define Λ_x and Λ_y as in section 3. We let $\{\Lambda_z\}$ be a partition of Ω and at present impose no further restrictions on the form of Λ_z (as for example is done in section 3.3). Then we partition Ω as

$$\{\Lambda_x^i\}_{i \in I} = \left\{ \bigcup_{w_X \in W_X} \Lambda(e(w_X, w_X^i)) \right\}_{i \in I}$$

where the edge $e(w_X, w_X^i)$ is the edge leaving w_X labelled x_i .

Substituting into the two conditional independence conditions we get the following

$$\begin{aligned} \pi(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_z) &= \pi(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^i, \Lambda_z) \\ &= \pi(\Lambda_y \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda(e(w_X, w_X^i)), \Lambda_z) \\ &= \pi(\Lambda_y \mid \Lambda(e(w_{X(1)}, w_{X(1)}^i)), \Lambda_z) \end{aligned} \quad (4.1)$$

and

$$\pi(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^i, \Lambda_z) = \pi(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^j, \Lambda_z)$$

for any $i, j \in I$. Also

$$\begin{aligned} \pi(\Lambda_z \mid \Lambda_x^i) &= \pi(\Lambda_z \mid \Lambda(w_{X(1)}), \Lambda_x^i) \\ &= \pi(\Lambda_z \mid \Lambda(w_{X(1)}), \bigcup_{w_X \in W_X} \Lambda(e(w_X, w_X^i))) \\ &= \pi(\Lambda_z \mid \Lambda(e(w_{X(1)}, w_{X(1)}^i))) \end{aligned} \quad (4.2)$$

and

$$\pi(\Lambda_z \mid \Lambda(w_{X(1)}), \Lambda_x^i) = \pi(\Lambda_z \mid \Lambda(w_{X(2)}), \Lambda_x^i)$$

for any $w_{X(1)}, w_{X(2)} \in W_X$ and any $i \in I$.

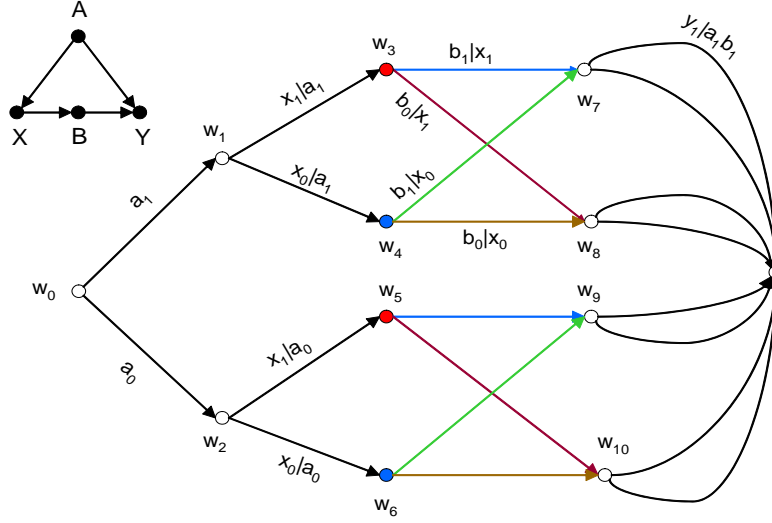


Figure 5: BN and CEG \mathcal{C} for Example 3

4.1. A Front Door theorem for singular manipulations

Theorem 2. (Front Door theorem) *If $\{\Lambda_x\}$ is defined as above, Λ_y is defined as in section 3, and $\{\Lambda_z\}$ is a partition of the $w_0 \rightarrow w_\infty$ paths in \mathcal{C} which satisfies conditions (4.1) and (4.2) above, then $\{\Lambda_z\}$ is a Front Door partition, and*

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_i \pi(\Lambda_y \mid \Lambda_x^i, \Lambda_z) \pi(\Lambda_x^i)$$

A proof of this theorem is in the appendix.

Note that unlike Pearl's Front Door theorem for BNs, Theorem 2 does not require the blocking set $\{\Lambda_z\}$ to lie downstream of the manipulation. This is clearly very useful.

Example 3. Using the Front Door theorem

We here consider the example from [9] section 3.3.3, but without reference to Pearl's hypothetical data. This example relates to the debate concerning the relationship between smoking and lung cancer summarised in [18].

In Pearl's example the vertices of the BN in Figure 5 correspond to binary variables as follows:

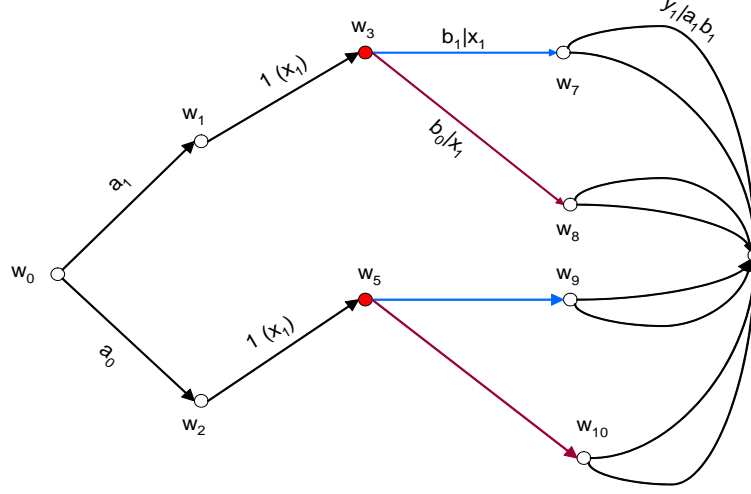


Figure 6: Manipulated CEG $\hat{\mathcal{C}}^{\Lambda_x}$ for Example 3

$X = 1$: smoker, $X = 0$: non-smoker,
 $Y = 1$: lung cancer, $Y = 0$: no lung cancer,
 $B = 1$: tar in lungs, $B = 0$: no tar in lungs.

The variable A is associated with an unobservable genetic tendency, the presence of which ($A = 1$) in an individual effects both the probability that the individual smokes and that they get lung cancer. The variable B by contrast is observable. Pearl uses the BN to show that it is possible to estimate $p(\text{lung cancer} \parallel \text{smoker})$ from joint or conditional distributions of the variables X, B and Y even if there were to exist such an unobservable genetic tendency.

We demonstrate the use of the Front Door theorem for CEGs by replicating this result. The unmanipulated CEG is given in Figure 5, where as before edges are labelled a_0 for $A = 0$ etc. We consider the manipulation to Λ_x equivalent to $Do X = 1$ and use Theorem 2 to find an expression for $\pi(\Lambda_y \parallel \Lambda_x) \equiv P(Y = 1 \parallel X = 1)$. The manipulated CEG $\hat{\mathcal{C}}^{\Lambda_x}$ is given in Figure 6.

Note that if A was observable we could use the Back Door theorem for CEGs here with $W_X = \{w_1, w_2\}$ doubling up as the blocking set (as in

section 3.4), which would be possible since each element of W_X is a distinct stage. Doing this we would get

$$\begin{aligned} p(y_1 \mid x_1) &= \hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_{u_X} \pi(\Lambda_y \mid \Lambda(u_X), \Lambda_x) \pi(\Lambda(u_X)) \\ &= \sum_a p(y_1 \mid a, x_1) p(a) \end{aligned}$$

For our Front Door theorem we have W_X as above, $\{\Lambda_x^i\} = \{\Lambda_x^1, \Lambda_x^2\}$, where

$$\begin{aligned} \Lambda_x^1 (= x_1) &= \Lambda(e(w_1, w_3)) \cup \Lambda(e(w_2, w_5)) \\ \Lambda_x^2 (= x_0) &= \Lambda(e(w_1, w_4)) \cup \Lambda(e(w_2, w_6)) \end{aligned}$$

The event Λ_y is expressible as $\bigcup_{w_Y} \Lambda(e^u(w_Y, w_\infty))$, where our $\{w_Y\}$ are $\{w_7, w_8, w_9, w_{10}\}$, and $e^u(w_Y, w_\infty)$ is the (upper) edge from w_Y to w_∞ labelled y_1 .

We here use the flexibility of CEG analysis to give each Λ_z a slightly different form from that used in Section 3.3. We use a form similar to that of Λ_x^i or Λ_y , and let

$$\Lambda_z^1 = \bigcup_{w_Z} \Lambda(e(w_Z, w_Z^1))$$

where our $\{w_Z\}$ are $\{w_3, w_4, w_5, w_6\}$, and the set $\{w_Z^1\}$ consists of w_7 corresponding to $w_Z = w_3, w_4$, and w_9 corresponding to $w_Z = w_5, w_6$.

Λ_z^2 is defined similarly, with $\{w_Z^2\}$ consisting of w_8 corresponding to $w_Z = w_3, w_4$, and w_{10} corresponding to $w_Z = w_5, w_6$.

Using a similar process to that utilised in the previous example, we can use the CEG \mathcal{C} to check very quickly that our partitions satisfy conditions (4.1) and (4.2).

Paths which are elements of $\Lambda(w_{X(1)}) \cap \Lambda_x^i \cap \Lambda_z^1$ pass through w_1 and w_7 for both $i = 1, 2$. The form of Λ_y and the result of Lemma 1 then imply that

$$\pi(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^1, \Lambda_z^1) = \pi(\Lambda_y \mid \Lambda(w_{X(1)}), \Lambda_x^2, \Lambda_z^1)$$

Similar results hold for $\Lambda(w_{X(1)}), \Lambda_z^2$; $\Lambda(w_{X(2)}), \Lambda_z^1$ and $\Lambda(w_{X(2)}), \Lambda_z^2$; and hence (4.1) holds.

The probability $\pi(\Lambda_z^1 \mid \Lambda(w_{X(1)}), \Lambda_x^1)$ is the probability $\pi_e(w_7 \mid w_3)$. But the positions w_3 and w_5 are in the same stage, and $\pi_e(w_7 \mid w_3) = \pi_e(w_9 \mid w_5) = \pi(\Lambda_z^1 \mid \Lambda(w_{X(2)}), \Lambda_x^1)$. Similar results hold for Λ_z^1, Λ_x^2 ; Λ_z^2, Λ_x^1 and Λ_z^2, Λ_x^2 ; and hence (4.2) holds.

Conditions (4.1) and (4.2) having been satisfied, we can substitute from the graph \mathcal{C} into the expression from Theorem 2 to get the Front Door expression for this example. Substituting $\Lambda_x^1 \equiv x_1$, $\Lambda_x^2 \equiv x_0$, $\Lambda_y \equiv y_1$, $\Lambda_z^1 \equiv b_1$ and $\Lambda_z^2 \equiv b_0$ into

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_z \pi(\Lambda_z \mid \Lambda_x^1) \sum_i \pi(\Lambda_y \mid \Lambda_x^i, \Lambda_z) \pi(\Lambda_x^i)$$

we get

$$p(y_1 \parallel x_1) = \sum_b p(b \mid x_1) \sum_x p(y_1 \mid x, b) p(x)$$

So as Pearl found, the expression $p(\text{lung cancer} \parallel \text{smoker})$ can be estimated from joint or conditional distributions of the variables X (smoker), B (tar in lungs) and Y (lung cancer) only.

4.2. A more flexible form of the Front Door theorem

At the start of section 4 we produced a partition of Ω of which Λ_x was **one** element, and noted that this meant confining ourselves to manipulations directly analogous to Pearl's *Do* $X = x$. This also required us to consider positions $\{w_X\}$ which had the same number of emanating edges and where these edges carried the same label for each w_X . In fact none of these restrictions is necessary, as we show here. One straightforward proof of Pearl's **Back** Door theorem proceeds as follows:

$$\begin{aligned} p(y \parallel x) &= \sum_{q(x), z} \left[\frac{p(q(x), x, z, y)}{p(x \mid q(x))} \right] \\ &= \sum_{q(x), z} p(q(x)) p(z \mid q(x), x) p(y \mid q(x), x, z) \end{aligned} \quad (4.3)$$

and then uses the conditional independence statements $Y \perp\!\!\!\perp Q(X) \mid (X, Z)$ and $Z \perp\!\!\!\perp X \mid Q(X)$ to remove $q(x)$ from (4.3) and leave the expression quoted at the start of Section 3.

Suppose instead we were to invoke the statements $Y \perp\!\!\!\perp X \mid (Q(X), Z)$ and $Z \perp\!\!\!\perp Q(X) \mid X$ when we reached expression (4.3). This would yield

$$p(y \parallel x) = \sum_z p(z \mid x) \sum_{q(x)} p(y \mid q(x), z) p(q(x))$$

This expression does not require knowledge of any joint probability including values of x other than the one to which we are manipulating. This leads to the following Corollary.

Corollary 1. *If W_X , Λ_x , Λ_y are defined as in section 3, and $\{\Lambda_z\}$ is a partition of the $w_0 \rightarrow w_\infty$ paths in \mathcal{C} which satisfies conditions (4.1) and (4.2), then $\{\Lambda_z\}$ is a Front Door partition, and*

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_z) \pi(\Lambda(w_X))$$

The proof of this corollary follows the proof of Theorem 2 until line (A.2).

This version of the Front Door theorem has a number of advantages over that given in Theorem 2, and over the Front Door theorem for BNs. Firstly we need to calculate or estimate a smaller number of joint probabilities than is the case with Theorem 2 (or the BN version which is an analogue of Theorem 2). This latter version is also appropriate, like the Back Door theorem of section 3.2, for the full range of singular manipulations, including both the *Do* $X = x$ and functional manipulations of BNs.

Note that like our Back Door theorem, both versions of the Front Door theorem for CEGs are suited for the analysis of asymmetric controlled models, and the Theorem 2 version allows us to use asymmetric blocking sets. The advantages of being able to do this are detailed in section 3.4. The Corollary 1 version allows us to analyse the effects of asymmetric manipulations, a task for which the Front Door theorem for BNs is manifestly unsuited.

5. Discussion

As noted in the Introduction, there have been a number of recent advances in BN theory which concentrate on the representation and analysis of asymmetric problems, and on the analysis of controlled models. The CEG is presented here as a complementary graphical model, appropriate for analysis in both these areas.

In this paper, the Back Door theorem of [22] has been generalised, and a Front Door theorem introduced. These theorems exhibit the flexibility of the CEG framework. They can both be used with all singular manipulations including the basic *Do* $X = x$ and functional manipulations possible on BNs. The Front Door theorem allows blocking sets which, unlike Pearl's for BNs, do not need to lie downstream of the manipulation. We have also provided a

version of the Front Door theorem which (again unlike the BN version) does not require us to sum over all values of a manipulated variable X .

Causal CEG analysis is still in its infancy. One potential direction for future investigation is the CEG's flexibility. So for example we have considered a partition $\{\Lambda_z\}$ which is *fixed* in the sense that its membership is constant. Causal analysis on CEGs would become even more flexible if we could let the membership of $\{\Lambda_z\}$ depend in some way on whichever $w_X \in W_X$ our $w_0 \rightarrow w_\infty$ path passes through. Looking at our Back Door theorem, the problem here would be in interpreting and satisfying condition (3.2), and it might prove more sensible to return to the original conditions (A) and (B) of Definition 5, rather than try to adapt conditions (3.1) and (3.2) to fit this situation. It would also be useful to adapt our Back and Front Door theorems to produce workable versions for some of the non-singular manipulations of the type described in [22] section 3.2.

Longer term, we aim to replicate the work of [3, 11, 23, 24] for BNs in producing necessary and sufficient conditions for causal identifiability, expressed as functions of the topology of the unmanipulated CEG.

Appendix A.

Proof of Theorem 1:

$$\hat{\pi}^{\Lambda_x}(\Lambda_y) = \sum_{w_X \in W_X} \hat{\pi}^{\Lambda_x}(\Lambda(w_X), \Lambda_y) = \sum_{w_X \in W_X} \hat{\pi}^{\Lambda_x}(\Lambda(w_X)) \hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda(w_X))$$

since $\{\Lambda(w_X)\}$ form a partition of the atomic events

$$= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda(w_X))$$

since every w_X lies upstream of our manipulation (Definition 6 (iv))

$$= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda(w_X), \Lambda(w'_X))$$

since $\Lambda(w_X) = \Lambda(e(w_X, w'_X)) \subset \Lambda(w'_X)$ in $\hat{\mathcal{C}}^{\Lambda_x}$

$$= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda(w'_X))$$

using the form specified for Λ_y , the fact that $w_X \prec w'_X \prec w_Y$ for some $w_Y \in W_Y$ in $\hat{\mathcal{C}}^{\Lambda_x}$, and the result of Lemma 1.

From the definition of our manipulation, any edge lying on a $w'_X \rightarrow w_\infty$ path in \mathcal{C} remains in $\hat{\mathcal{C}}^{\Lambda_x}$, and retains its original probability. Hence any set of path-segments starting at w'_X in $\hat{\mathcal{C}}^{\Lambda_x}$ corresponds to a set of path-segments in \mathcal{C} , and has the same probability as this set. Given the form specified for Λ_y , $\hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda(w'_X))$ is the probability of a set of path-segments starting at w'_X in $\hat{\mathcal{C}}^{\Lambda_x}$. Hence

$$\hat{\pi}^{\Lambda_x}(\Lambda_y \mid \Lambda(w'_X)) = \pi(\Lambda_y \mid \Lambda(w'_X))$$

and

$$\begin{aligned} \hat{\pi}^{\Lambda_x}(\Lambda_y) &= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \pi(\Lambda_y \mid \Lambda(w'_X)) \\ &= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \pi(\Lambda_y \mid \Lambda(e(w_X, w'_X)), \Lambda(w'_X)) \end{aligned}$$

using the form specified for Λ_y , the fact that $e(w_X, w'_X) \prec w'_X \prec w_Y$ for some $w_Y \in W_Y$ in \mathcal{C} , and the result of Lemma 1

$$= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \pi(\Lambda_y \mid \Lambda(e(w_X, w'_X)))$$

since $\Lambda(e(w_X, w'_X)) \subset \Lambda(w'_X)$ in \mathcal{C}

$$= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \sum_z \pi(\Lambda_z, \Lambda_y \mid \Lambda(e(w_X, w'_X)))$$

since $\{\Lambda_z\}$ form a partition of the atomic events

$$\begin{aligned} &= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \sum_z \pi(\Lambda_y \mid \Lambda(e(w_X, w'_X)), \Lambda_z) \\ &\quad \times \pi(\Lambda_z \mid \Lambda(e(w_X, w'_X))) \tag{A.1} \\ &= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z) \pi(\Lambda_z \mid \Lambda(w_X)) \end{aligned}$$

substituting from (3.1) and (3.2)

$$= \sum_z \pi(\Lambda_y \mid \Lambda_x, \Lambda_z) \pi(\Lambda_z) \quad \square$$

Proof of Theorem 2:

This follows the proof of Theorem 1 until line (A.1). We then invoke conditions (4.1) and (4.2) to give

$$\begin{aligned}
\hat{\pi}^{\Lambda_x}(\Lambda_y) &= \sum_{w_X \in W_X} \pi(\Lambda(w_X)) \sum_z \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_z) \pi(\Lambda_z \mid \Lambda_x) \\
&= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_z) \pi(\Lambda(w_X)) \\
&= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \sum_i \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_z) \pi(\Lambda(w_X), \Lambda_x^i)
\end{aligned} \tag{A.2}$$

since $\{\Lambda_x^i\}$ forms a partition of Ω

$$= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \sum_i \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_x^i, \Lambda_z) \pi(\Lambda(w_X), \Lambda_x^i)$$

using condition (4.1). But

$$\pi(\Lambda(w_X), \Lambda_x^i) = \frac{\pi(\Lambda(w_X), \Lambda_x^i, \Lambda_z)}{\pi(\Lambda_z \mid \Lambda(w_X), \Lambda_x^i)} = \frac{\pi(\Lambda(w_X), \Lambda_x^i, \Lambda_z)}{\pi(\Lambda_z \mid \Lambda_x^i)}$$

using condition (4.2)

$$= \pi(\Lambda(w_X) \mid \Lambda_x^i, \Lambda_z) \pi(\Lambda_x^i)$$

So

$$\begin{aligned}
\hat{\pi}^{\Lambda_x}(\Lambda_y) &= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_{w_X \in W_X} \sum_i \pi(\Lambda_y \mid \Lambda(w_X), \Lambda_x^i, \Lambda_z) \\
&\quad \times \pi(\Lambda(w_X) \mid \Lambda_x^i, \Lambda_z) \pi(\Lambda_x^i) \\
&= \sum_z \pi(\Lambda_z \mid \Lambda_x) \sum_i \pi(\Lambda_y \mid \Lambda_x^i, \Lambda_z) \pi(\Lambda_x^i) \quad \square
\end{aligned}$$

Acknowledgements

This research is being supported by the EPSRC, grant no. EP/F036752/1.

References

- [1] C. Boutilier, N. Friedman, M. Goldszmidt, D. Koller, Context-specific independence in Bayesian Networks, in: Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, 1996, pp. 115–123.
- [2] A.P. Dawid, Influence diagrams for causal modelling and inference, *International Statistical Review* 70 (2002) 161–89.
- [3] A.P. Dawid, V. Didelez, Identifying the consequences of dynamic treatment strategies, Research Report 262, University College London, 2005.
- [4] D. Glymour, G.F. Cooper, *Computation, Causation and Discovery*, MIT Press, 1999.
- [5] D. Heckerman, A Bayesian approach to Learning Causal Networks, in: W. Edwards, et al. (Eds.), *Advances in Decision Analysis*, CUP, 2007, pp. 202–220.
- [6] S.L. Lauritzen, Causal inference from graphical models, in: O.E. Barndorff-Nielsen, et al. (Eds.), *Complex Stochastic Systems*, Chapman and Hall, 2001.
- [7] D. McAllester, M. Collins, F. Periera, Case factor diagrams for structured probabilistic modeling, in: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, 2004, pp. 382–391.
- [8] J. Pearl, Causal diagrams for empirical research, *Biometrika* 82 (1995) 669–710.
- [9] J. Pearl, *Causality: Models, Reasoning and Inference*, Cambridge, 2000.
- [10] J. Pearl, Statistics and causal inference: A review, *Sociedad de Estadística e Investigación Operativa. Test* 12 (2003) 281–345.
- [11] J. Pearl, J.M. Robins, Probabilistic evaluation of sequential plans from causal models with hidden variables, in: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, 1995, pp. 444–45.
- [12] D. Poole, N.L. Zhang, Exploiting contextual independence in probabilistic inference, *Journal of Artificial Intelligence Research* 18 (2003) 263–313.

- [13] J.M. Robins, A new approach to causal inference in mortality studies with sustained exposure period – application to control of the healthy worker survivor effect, *Mathematical Modelling* 7 (1986) 1393–1512.
- [14] A. Salmeron, A. Cano, S. Moral, Importance sampling in Bayesian Networks using probability trees, *Computational Statistics and Data Analysis* 34 (2000) 387–413.
- [15] G. Shafer, *The Art of Causal Conjecture*, MIT Press, 1996.
- [16] J.Q. Smith, P.E. Anderson, Conditional independence and Chain Event Graphs, *Artificial Intelligence* 172 (2008) 42–68.
- [17] J.Q. Smith, P.A. Thwaites, Influence diagrams, in: E.L. Melnick, B.S. Everitt (Eds.), *Encyclopedia of Quantitative Risk Analysis and Assessment*, volume 2, Wiley, 2008, pp. 897–910.
- [18] P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction and Search*, Springer-Verlag, 1993.
- [19] P.A. Thwaites, *Chain Event Graphs: Theory and application*, Ph.D. thesis, University of Warwick, 2008.
- [20] P.A. Thwaites, J.Q. Smith, Separation theorems for Chain Event Graphs, Submitted to the *Annals of Statistics* (2011).
- [21] P.A. Thwaites, J.Q. Smith, R.G. Cowell, Propagation using Chain Event Graphs, in: *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, 2008, Helsinki, pp. 546–553.
- [22] P.A. Thwaites, J.Q. Smith, E.M. Riccomagno, Causal analysis with Chain Event Graphs, *Artificial Intelligence* 174 (2010) 889–909.
- [23] J. Tian, Identifying dynamic sequential plans, in: *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, 2008, Helsinki, pp. 554–561.
- [24] J. Tian, J. Pearl, A general identification condition for causal effects, in: *Proceedings of the 18th National Conference on Artificial Intelligence*, AAAI Press, 2002, pp. 567–573.